

DIAGNOSA PENYAKIT GINJAL BEDASARKAN DATA PASIEN ICU DENGAN ALGORITMA CATBOOST

Shalsabilillah Naufal

Pasca Sarjana Teknik Informatika, Teknologi Informatika , Universitas Yarsi, DKI Jakarta

Shalsabilillah.naufal@gmail.com

Abstract : Kidney disease is a type of disease that causes a gradual decline in kidney function. The prevalence of kidney disease increases as the elderly population increases. According to the 2010 Global Burden of Disease, CKD was the 27th leading cause of death in the world in 1990 and increased to 18th in 2010. Meanwhile, in Indonesia, treatment of kidney disease is the second largest health insurance provider after heart disease. This study aims to implement machine learning methods to detect the risk of kidney disease, make it easier for health workers to shorten the time for diagnosing kidney disease, especially in ICU patients, and provide an Islamic review of the diagnosis of kidney disease. In this study, the authors preprocessed with data cleaning and fill missing values with the fillna technique, did a data split of 90% for train data and 10% for train data, and compared 3 algorithms, namely: Random Forest, Decision Tree, and CatBoost. This study only focuses on the AUC score to compare each algorithm and data on unbalanced research targets. With the method used by the CatBoost algorithm, the highest AUC score is 92.21%, while for Random Forest it is 89.28%, and Decision Tree is 70.23%.

Keyword : Machine Learning, Kidney Disease, preprocessing, Missing Value, CatBoost

Abstrak : Penyakit ginjal merupakan jenis penyakit yang menyebabkan penurunan fungsi ginjal secara bertahap. Prevalansi penyakit ginjal meningkat seiring naiknya penduduk usia lanjut. Menurut Global Burden of Disease tahun 2010, PGK merupakan penyebab kematian peringkat ke-27 di dunia tahun 1990 dan meningkat menjadi urutan ke-18 pada tahun 2010. Sedangkan di Indonesia, perawatan penyakit ginjal merupakan ranking kedua pembiayaan terbesar dari BPJS Kesehatan setelah penyakit jantung. Penelitian ini bertujuan untuk mengimplementasikan metode machine learning untuk mendeteksi resiko penyakit ginjal, memudahkan tenaga Kesehatan untuk mempersingkat waktu diagnosa penyakit ginjal khususnya pada pasien ICU, dan memberikan tinjauan islam terhadap diagnosa penyakit ginjal. Pada penelitian ini penulis melakukan preprocessing dengan data cleaning dan fill missing value dengan teknik fillna, melakukan data split 90% untuk data train dan 10% untuk data train, dan membandingkan 3 algoritma yaitu : Random Forest, Decision Tree, dan CatBoost. Penelitian ini hanya berfokus pada AUC score untuk membandingkan setiap algoritma dan data pada target penelitian yang tidak seimbang. Dengan metode yang dilakukan algoritma CatBoost mendapatkan AUC score tertinggi sebesar 92.21%, sedangkan untuk Random Forest sebesar 89.28%, dan Decision Tree sebesar 70.23%.

Kata Kunci : Machine Learning, Penyakit Ginjal, Preprocessing, Missing Value, CatBoost

1. Pendahuluan

Penyakit Ginjal Kronis Merupakan jenis penyakit yang menyebabkan penurunan fungsi ginjal secara bertahap. Prevalansi penyakit ginjal meningkat seiring meningkatnya penduduk usia lanjut. Menurut Global Burden of Disease tahun 2010, penyebab kematian PGK menduduki peringkat ke-27 pada tahun 1990 dan meningkat menjadi ke -18 pada tahun

2010. Sedangkan di Indonesia, perawatan PGK merupakan ranking kedua pembiayaan terbesar dai BPJS Kesehatan setelah penyakit jantung. Banyaknya pasien Penyakit Ginjal Kronis yang tidak patuh terhadap pola diet atau anjuran dari tenaga kesetahan.

Pada pasien PGK yang menjalani HD (Hemodialisa) rutin sering kali mengalami kelebihan volume cairan yang terdapat pada tubuh ini disebabkan karena penurunan fungsi ginjal. Faktor-faktor yang mempengaruhi kepatuhan dalam mengurangi asupan cairan dan memberikan asuhan keperawatan, diantaranya factor pendidikan, keterlibatan keluarga, dan juga keterlibatan tenaga Kesehatan yang bertugas untuk memotivasi pasien supaya termotivasi untuk mengikuti aturan atau anjuran. (Kamaluddin & Rahayu, 2009).

Banyak pasien penyakit ginjal kronis tidak patuh terhadap pola diet. Pasien harus patuh dengan pola diet sesuai anjuran yang disarankan. Kurangnya kepatuhan pasien disebabkan karena kurangnya pemahaman terhadap anjuran yang diberikan dari tenaga Kesehatan. Ketidak patuhan pasien memberikan dampak defisiensi gizi, tidak teraturnya keseimbangan cairan dan elektrolit akan terganggu yang menyebabkan akan terjadinya peningkatan sisa metabolisme (uremia) yang berlebihan (Zendrato et al., 2015).

Berdasarkan permasalahan diatas, penulis melakukan diagnose pada pasien ICU dengan menggunakan pendekatan algoritma Catboost klasifikasi dari metode machine learning. Untuk mendapatkan prediksi yang optimal, penulis melakukan data cleaning untuk menghilangkan informasi yang tidak penting dan missing value dengan Teknik *fillna*. Serta data dibagi menjadi 2 bagian, yaitu data train dan test.

2. Landasan Teori

2.1 Penyakit Ginjal Kronis

Penyakit ginjal Kronis merupakan jenis penyakit yang menyebabkan penurunan fungsi ginjal secara bertahap. Pasien penyakit ginjal biasanya mengalami kelebihan volume cairan yang terdapat pada tubuh. Kelebihan volume cairan disebabkan oleh kurangnya kepatuhan pasien dalam menjalani diet atau anjuran dari tenaga Kesehatan (Kamaluddin & Rahayu, 2009).

Kepatuhan merupakan kedisiplinan pasien untuk menjalankan diet atau melakukan terapi yang dianjurkan. Kurangnya kepatuhan pada pasien disebabkan karena kurangnya pemahaman pasien terhadap arahan dari tenaga kesehatan (Zendrato et al., 2015).

2.2 Machine Learning

Machine Learning merupakan bagian dari Artificial Intelligence (AI) yang berfungsi untuk menirukan proses atau perilaku yang dilakukan manusia untuk menyelesaikan masalah atau melakukan perhitungan. Ada dua jenis penerapan machine learning yang digunakan secara umum, yaitu klasifikasi dan prediksi. Klasifikasi merupakan metode machine learning yang berfungsi untuk mengelompokkan atau memilah suatu objek berdasarkan objek tertentu. Sedangkan prediksi berfungsi untuk menebak hasil output dari data yang sudah dipelajari.

2.3 Preprocessing

Preprocessing merupakan teknik yang digunakan untuk mengubah data mentah menjadi data yang siap untuk diolah. Kualitas data berpengaruh terhadap keberhasilan suatu proyek.

Secara umum tahapan preprocessing ada dua, yaitu data cleaning dan missing value. Data cleaning merupakan pemrosesan data untuk dilakukan analisis dengan cara menghapus atau merubah data yang salah. Missing value merupakan informasi yang tidak diketahui nilainya.

2.4 Algoritma CatBoost

CatBoost merupakan algoritma open-source yang dibuat oleh Yandex. Algoritma ini bagian dari Gradient Boosting yang digunakan untuk klasifikasi. CatBoost juga dapat mengurangi *overfitting* pada data untuk meningkatkan keakuratan performa model. (Lunden, 2017).

3. Metodologi

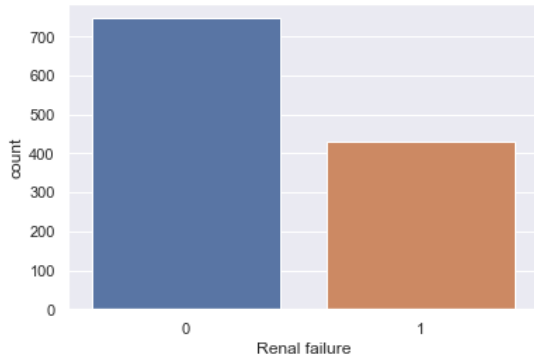
3.1 Data Penelitian

Tabel 1 Dataset In Hospital Mortality

group	ID	outcome	age	gender	BMI
1	125047	0	72	1	37.58818
1	139812	0	75	2	NA
1	109787	0	83	2	26.57263
1	130587	0	43	2	83.26463
1	138290	0	75	2	31.82484

Data yang digunakan pada penelitian ini adalah menggunakan data RS. Mortality pada pasien yang dirawat pada Intensive Care Units (ICU), data tersebut merupakan database (MIMIC-III versi 1.4) dan dapat diakses pada situs Kaggle¹. Dataset terdiri 1177 baris dan 51 kolom.

¹ <https://www.kaggle.com/saurabhshahane/in-hospital-mortality-prediction>

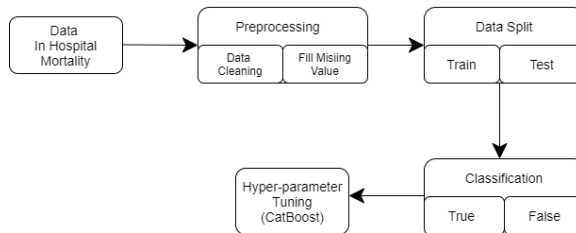


Gambar 1 Target Penelitian

Pada gambar diatas dapat dinyatakan bahwa dataset memiliki sampel tidak merata. Jumlah sampel pasien yang terdiagnosa penyakit ginjal lebih sedikit yaitu sebanyak 430 pasien dan pasien yang tidak terdiagnosa penyakit ginjal sebanyak 747 pasien

3.2 Data Eksperimen

Alur Program



Gambar 2 Alur Program

Tahap pertama dalam penelitian ini adalah mendapatkan dataset In Hospital Mortality yang terdapat dari Kaggle. Data lalu dilakukan preprocessing dengan cara data cleaning dan fill missing value dengan teknik fillna. Kemudian data dilakukan data split untuk membagi data. Dan terakhir data dilakukan uji algoritma.

3.3 Preprocessing

Preprocessing dilakukan untuk merapihkan bentuk data dan dapat diproses agar sesuai dengan tujuan pemrosesan yang diinginkan. Pada penelitian ini penulis melakukan 2 cara pada preprocessing, yaitu data cleaning dan fill missing value.

Dari data In Hospital Mortality terdapat kolom 'ID' merupakan sebuah indeks, informasi ini tidak diperlukan dalam melakukan diagnose penyakit ginjal. Maka kolom 'ID' dapat dihapus.

Penulis menggunakan fill missing value dengan teknik fillna. Teknik ini dilakukan untuk

memperhitungkan nilai yang hilang dari satu atau lebih kolom fitur.

3.4 CatBoost

CatBoost merupakan model machine learning yang andal, skalabel, dan Tangguh yang memungkinkan untuk meningkatkan kinerja berdasarkan system peningkatan gradien dan gradien dan decision trees secara bersamaan. CatBoost dapat mengurangi overhead data dari tipe data kategori menjadi numerik dengan mudah (JournalDev).

3.5 Evaluasi Model

Evaluasi yang digunakan pada penelitian ini adalah menghitung nilai akurasi, AUC, Precision, Recall, dan F1 Score. Pada penelitian ini penulis hanya berfokus dengan hasil AUC Score dari setiap algoritma. AUC Score digunakan karena tidak seimbangnya jumlah pasien yang terdiagnosa dan yang tidak terdiagnosa. Kriteria keakuratan tes diagnostik menggunakan AUC disajikan pada tabel (Gorunescu, 2011).

Tabel 2 Kriteria nilai AUC

Nilai AUC	Interpretasi
0.90 – 1.00	<i>Excellent classification</i>
0.80 – 0.90	<i>Good classification</i>
0.70 – 0.80	<i>Fair classification</i>
0.60 – 0.70	<i>Poor classification</i>
0.50 – 0.60	<i>Failure</i>

Tabel 3 merupakan acuan penulis dalam menilai hasil AUC yang ada pada penelitian ini. Kalsifikasi dikatakan bagus jika AUC Score diatas 0.80 dan dikategorikan buruk ketika dibawah 0.80.

4. Hasil dan Pembahasan

4.1 Preprocessing

Pada penelitian ini penulis melakukan data cleaning dan fill missing value dengan teknik fillna. Untuk data cleaning penulis menghapus 1 kolom yaitu kolom 'ID', sehingga tersisa 50 kolom.

group	ID	outcome	age	gender	BMI	hypertensive	atrialfibrillation
1	125047	0.0	72	1	37.588179	0	0
1	139812	0.0	75	2	NaN	0	0
1	109787	0.0	83	2	26.572634	0	0
1	130587	0.0	43	2	83.264629	0	0
1	138290	0.0	75	2	31.824842	1	0

(a)

	group	outcome	age	gender	BMI	hypertensive	atrialfibrillation
0	1	0.0	72	1	37.588179	0	0
1	1	0.0	75	2	NaN	0	0
2	1	0.0	83	2	26.572634	0	0
3	1	0.0	43	2	83.264629	0	0
4	1	0.0	75	2	31.824842	1	0

(b)

Gambar 3 Hasil Data Cleaning

Gambar 3.a merupakan data sebelum dilakukan data cleaning sedangkan gambar 3.b merupakan data setelah dilakukan data cleaning.

Dari dataset yang digunakan terdapat 20 kolom, terdapat beberapa baris yang tidak terisi atau kosong. Penulis melakukan fill missing value dengan teknik fillna, yaitu dengan mengisi data berdasarkan nilai mean.

BMI	BMI
37.588179	37.588179
NaN	30.188278
26.572634	26.572634
83.264629	83.264629
31.824842	31.824842

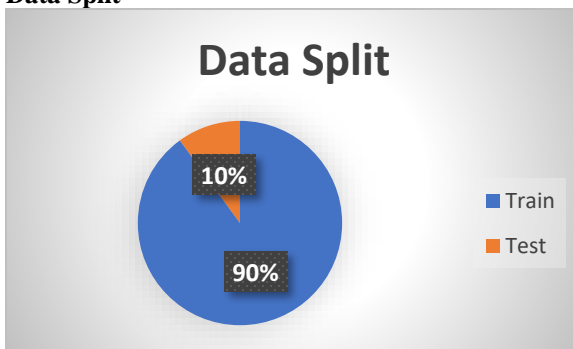
(a)

(b)

Gambar 4 Hasil Fill Missing Value

Gambar 4.a merupakan data sebelum dilakukan fill missing value dengan teknik fillna sedangkan gambar 4.b merupakan data sesudah dilakukan fill missing value dengan teknik fillna.

4.2 Data Split



Gambar 5 Pembagian Data Split

Penulis membagi data dengan perbandingan 90:10, 90% untuk data train dan 10% data test. Hasil pembagian dengan menggunakan metode ini adalah 1059 baris untuk data train dan 118 baris untuk data test.

4.3 Hasil Uji Algoritma CatBoost

Classification with CatBoost
[1 1 0 1 1 1 0 0 0 1 0 1 1 1 1]

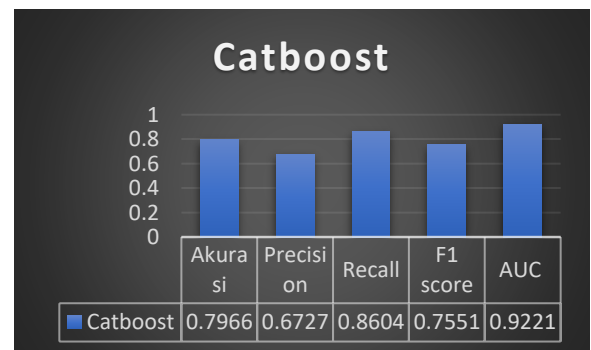
(a)

Renal failure
1
0
1
0
1
1
1
0
1
0
0
0
0
1
0

(b)

Gambar 6 Klasifikasi Penyakit Ginjal

Gambar 6.a merupakan hasil klasifikasi atau diagnosa menggunakan algoritma CatBoost dan gambar 6.b merupakan data asli in hospital mortality. Berdasarkan gambar tersebut algoritma catboost mempunyai akurasi yang baik.

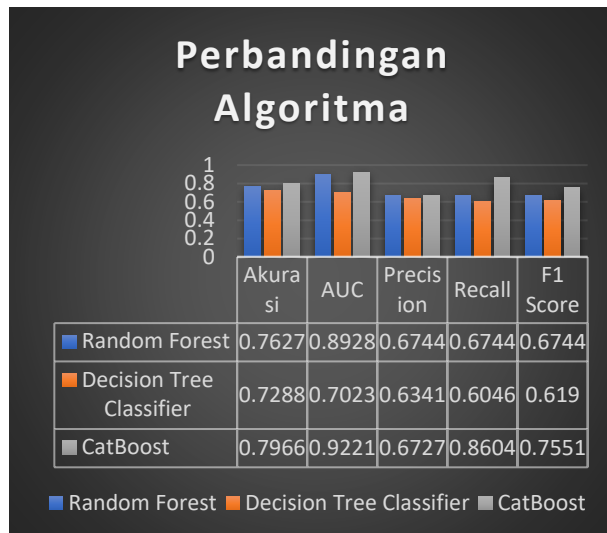


Gambar 7 Hasil Algoritma CatBoost

Gambar diatas merupakan hasil dari algoritma CatBoost. Setelah melakukan percobaan algoritma ini menghasilkan akurasi 79.66%, Precision 67.27%, Recall 86.04%, F1 Score 75.51%, dan AUC Score 92.21%.

4.4 Evaluasi Model

Pada penelitian ini penulis menggunakan data in Hospital Mortality pada pasien yang dirawat di Intensive Care Units (ICU), data tersebut dapat di akses pada situs Kaggle. Dataset memiliki 1177 baris dan 51 kolom.



Gambar 7 Hasil Evaluasi Perbandingan Algoritma

Berdasarkan gambar tersebut dan dengan metode yang sama dilakukan dengan cara data cleaning, fill missing value dengan teknik fillna, dan data split 90:10 algoritma CatBoost memiliki AUC Score tertinggi dibandingkan dengan algoritma Random Forest dan Decision Tree.

5. Tinjauan Islam

5.1 Penyakit Ginjal Menurut Tinjauan Islam

Islam mengedepankan tindakan pencegahan dibandingkan pengobatan. Dalam konteks penyakit ginjal, islam telah memerintahkan dalam surah Al-Baqarah ayat 168 yang berbunyi :

يَا أَيُّهَا النَّاسُ كُلُوا مِمَّا فِي الْأَرْضِ حَلَالًا طَيِّبًا وَلَا تَتَّبِعُوا خُطُوَاتِ الشَّيْطَانِ إِنَّهُ لَكُمْ عَدُوٌّ مُبِينٌ ١٦٨

(البقرة/2:168)

Artinya : “Wahai manusia, makanlah sebagian (makanan) di bumi yang halal lagi baik dan janganlah mengikuti langkah-langkah setan. Sesungguhnya ia bagimu merupakan musuh yang nyata.” (QS. Al-Baqarah[2]:168).

Manusia diperintahkan memakan makanan yang halal, yaitu yang diperbolehkan atau tidak diharamkan. Baik itu kandungan pada makanannya maupun cara

memperolehnya. Selain itu makanan juga harus sehat dan bermanfaat serta tidak dikonsumsi secara berlebihan.

5.2 Penerapan Ilmu Pengetahuan dan Teknologi

Ilmu pengetahuan dapat berkembang menghasilkan sebuah teknologi. Banyak ilmu yang dapat diambil dari membaca alam sekitar, dan bahkan dapat memperdalam iman dan keyakinan tentang pencipta alam semesta ini, yaitu Allah SWT (Raya, 2020). Sebagaimana Allah berfirman :

اقْرَأْ بِاسْمِ رَبِّكَ الَّذِي خَلَقَ ١ خَلَقَ الْإِنْسَانَ مِنْ عَلَقٍ ٢ اقْرَأْ وَرَبُّكَ الْأَكْرَمُ ٣ الَّذِي عَلَّمَ بِالْقَلَمِ ٤ عَلَّمَ الْإِنْسَانَ مَا لَمْ يَعْلَمْ ٥

(العلق/96:1-5)

Artinya : “Bacalah dengan (menyebut) nama Tuhanmu yang menciptakan!. Dia menciptakan manusia dari segumpal darah. Bacalah! Tuhanmulah Yang Mahamulia, yang mengajar (manusia) dengan pena. Dia mengajarkan manusia apa yang tidak diketahuinya.” (QS. Al-'Alaq[96]:1-5)

Perintah untuk membaca kepada Nabi Muhammad SAW menghasilkan bacaan yang banyak, dengan membaca seseorang mendapatkan banyak ilmu dan pengetahuan. (Raya, 2020).

5.3 Tinjauan Islam Pada Diagnosa Penyakit Ginjal Pada Pasien ICU

Menurut Kamus Besar Bahasa Indonesia (KBBI) Diagnosis adalah penentuan jenis penyakit dengan cara meneliti (memeriksa) gejala-gejalanya. Secara umum ramalan diartikan sebagai prediksi atau memperkirakan sesuatu yang akan terjadi dimasa mendatang dengan data, pendapat, atau perkiraan seseorang.

Menurut Dalamislam (*Hukum Percaya Ramalan Menurut Islam - Boleh Atau Tidak? - DalamIslam.Com*, n.d.) ramalan terdiri menjadi 2 bagian yaitu, ramalan ilmiah dan ramalan berdasarkan dari jin atau tanpa dasar.

Ramalan ilmiah merupakan ramalan yang berasal dari perkiraan yang berdasarkan ilmu pengetahuan. Kategori ramalan ini masih diperbolehkan selagi memiliki manfaat dan kemasalahatan bagi umat.

بَلْ اتَّبَعَ الَّذِينَ ظَلَمُوا أَهْوَاءَهُمْ بِغَيْرِ عِلْمٍ ۖ فَمَنْ يَهْدِي مَنْ أَضَلَّ اللَّهُ ۚ وَمَا لَهُمْ مِنْ نَاصِرِينَ

Artinya : Tetapi orang-orang yang zalim, mengikuti hawa nafsunya tanpa ilmu pengetahuan; maka

siapakah yang akan menunjuki orang yang telah disesatkan Allah? Dan tiadalah bagi mereka seorang penolongpun. QS. Ar-Rum [30] : 29)

Walaupun bersifat ilmiah dan berdasarkan pengetahuan ramalan jenis ini pun juga masih ada kemungkinan tidak tepat. Hal ini dikarenakan adanya ketidaktelitian, kurangnya variable yang diperkirakan, dan sebagainya. Oleh karena itu peramalan jenis ini bersifat perkiraan, reference, dan bukan untuk dijadikan kepercayaan yang mutlak sebagai yang paling benar.

5.4 Kesimpulan

Pada penelitian skripsi ini termasuk ke dalam ramalan ilmiah karena menggunakan data pasien *Intensive Care Unit* (ICU) pada in hospital mortality. Penelitian ini dapat dijadikan referensi agar pasien ICU dapat terdiagnosa penyakit ginjal dengan cepat. Sehingga diharapkan penelitian ini dapat membantu tenaga Kesehatan untuk dapat melakukan perawatan yang sesuai.

6. Penutup

6.1 Kesimpulan

Penelitian ini bertujuan untuk melakukan klasifikasi data pasien ICU berdasarkan data in hospital mortality. Tahapan yang dilakukan penulis untuk penelitian ini adalah preprocessing dengan melakukan data cleaning dan fill missing value dengan teknik fillna, data split dengan komposisi 90:10, uji model dan evaluasi model.

Bedasarkan eksperimen dan analisa yang dilakukan, diagnose penyakit ginjal berdasarkan data pasien ICU dengan menggunakan algoritma *CatBoost* mendapatkan hasil :

1. Metode *data cleaning*, *fill missing value* dengan teknik *fillna*, dan data split 90:10 dapat mempengaruhi hasil akhir diagnose penyakit ginjal dari klasifikasi data pasien ICU in hospital mortality.
2. Algoritma *CatBoost* mendapatkan skor AUC tertinggi dibandingkan algoritma *Decision Tree* dan *Random Forest*.
3. Menurut sudut pandang islam, penelitian ini termasuk dalam ramalan ilmiah karena merupakan perkiraan berdasarkan ilmu pengetahuan dan menggunakan data pasien ICU pada in hospital mortality.

6.2 Saran

Bedasarkan penelitian yang penulis lakukan, saran yang dapat penulis sampaikan untuk mengembangkan penelitian selanjutnya :

1. Menggunakan korelasi untuk melihat data yang terkait supaya melakukan data cleaning.
2. Menggunakan parameter *CatBoost* yang berbeda untuk menaikkan nilai akurasi.
3. Menggunakan algoritma *Machine Learning* lain.
4. Melakukan teknik pembagian data yang berbeda.

Daftar Pustaka

Al-Qur'an dan Terjemahan. Kementerian Agama Republik Indonesia. 2019.

<https://www.kaggle.com/saurabhshahane/in-hospital-mortality-prediction>

ALADOKTES (2018). Gagal Ginjal Kronis. <https://www.alodokter.com/gagal-ginjal-kronis>

halodoc (2019). Gagal Ginjal Kronis. <https://www.halodoc.com/kesehatan/gagal-ginjal-kronis>

vidyasheela. (2021). What Is Data Pre-Processing ?. <https://vidyasheela.medium.com/what-is-data-pre-processing-e3bef2fbf461>

vitalflux, Ajitesh Kumar. (2020). Pandas – Fillna method for replacing missing values. <https://vitalflux.com/pandas-fillna-method-imputing-missing-values/>

catboost.ai. (2018). CatBoost is a high-performance open source library for gradient boosting on decision trees. <https://catboost.ai/news/catboost-enables-fast-gradient-boosting-on-decision-trees-using-gpus>

CatBoost, "How training is performed", CatBoost, [online]. <https://catboost.ai/docs/concepts/algorithm-main-stages.html>

Rahayu, E., Kamaluddin, R. (2009). ANALISIS FAKTOR-FAKTOR YANG MEMPENGURHI KEPATUHAN ASUPAN CAIRAN PADA PASEIEN GAGAL GINJAL KRONIK DENGAN HEMODILISIS DI RSUD Prof.Dr MARGONO SOEKARJO PURWOKERTO. file:///D:/skripsi/Chronic%20Kidney%20Disease/jurn

al/PGK/Analisis%20Faktor-Faktor%20yang%20Mempengaruhi%20Kepatuhan%20Asupan%20Cairan%20pada%20Pasien%20Gagal%20Ginjal%20Kronik%20dengan%20Hemodialisis%20di%20RSUD%20Prof.%20Dr.%20Margono%20Soekarjo%20Purwokerto%20_%20Kamaluddin%20_%20Jurnal%20Keperawatan%20Soedirman.pdf

Zendrato, C. F., Susilawati, E., Taringan, A. P. S. (2015). FAKTOR-FAKTOR YANG MEMENGARUHI KEPATUHAN DIET PASIEN GAGAL GINJAL KRONIK YANG MENDAPAT TERAPI HEMODIALISA DI RUANG HEMODIALISA RSUD Dr. PIRNGADI KOTA MEDAN TAHUN 2015. <http://ojs.poltekkes-medan.ac.id/pannmed/article/view/114/101>

Safira, A. P. (2021). Apa Itu Machine Learning? Pengertian, Cara Kerja & Jenisnya. <https://www.goldenfast.net/blog/apa-itu-machine-learning/>

Ahmad, A. (2017). Mengenal Artificial Intelligence, Machine Learning, Neural Network, dan Deep Learning. https://d1wqtxts1xzle7.cloudfront.net/54674088/Perbedaan_Deep_learn-with-cover-page-v2.pdf?Expires=1637412546&Signature=B5c9BEXT1aPu1qincBoJ09xUVIdrOYAr08nCFyWyymboblw18hWyMVgOujuMWTqabP~V5sd58p-r5APsoUVQfvqUI2Ty8nj3A4-U1tKy6x3Xhgkg7Ym9dQSawXIvZF4Hv8fDObNqdHyXLkR19KaCnctq6RKAmr3rGnkJXqrFAXfE6nDSwH56O0LFpVs8CvU72g6HYTUfhp~93ntEzinQDMMYUSVdx1lsUpdi2EcAFEDcpqR9s9jIYI6AhvwPN9hOfFqnjEDUAoGDP0L4LAHWGqnJCX9GhTYOVv0l4jyIOHZN8tHQhIk8WnZPEuWXp5Mq0nBxWYBhI3cRtot1Nsn9Q__&Key-Pair-Id=APKAJLOHF5GGSLRBV4ZA

Samir. (2019). Apa itu Preprocessing. <https://hendroprasetyo.com/apa-itu-preprocessing/#.YZI1qtBBzIU>

Oliver, A. (2021). Bikin Data Lebih Mudah Dibaca, yuk, Kenalan dengan Data PreprocessingBikin Data Lebih Mudah Dibaca, yuk, Kenalan dengan Data Preprocessing. <https://glints.com/id/lowongan/data-preprocessing-adalah/#.YZoy7dBBzIV>

Mail. (2020). Pengertian Data Cleaning Penyebab data menjadi Unclean dan Teknik. <https://anaktik.com/data-cleaning/>

Khalimi, A. M. (2020). Algoritma C4.5 Pada Kasus Missing Value. <https://www.pengalaman-edukasi.com/2020/09/evaluasi-algoritma-c45-pada-kasus.html>

Gorunescu, F. (2011). *Data Mining: Concepts, Models and Techniques - Florin Gorunescu - Google Books*. https://books.google.co.id/books?hl=en&lr=&id=yJvKY-sB6zkC&oi=fnd&pg=PP2&dq=Data+mining:+concepts+and+techniques+gorunescu&ots=puJyu9wSAB&sig=Y_e7xXI7JFJXuPVAWs7ey2XUwng&redir_esc=y#v=onepage&q=roc&f=false

DalamIslam (*Hukum Percaya Ramalan Menurut Islam - Boleh Atau Tidak? - DalamIslam.Com*, n.d.)

Raya, A. T. (2020). *Tadabbur Atas Surat Al-Alaq Ayat 1-5: Wahyu Pertama Perintah Membaca*. <https://tafsiralquran.id/tadabbur-atas-surat-al-alaq-ayat-1-5-wahyu-pertama-perintah-membaca/>